

CLAIMS

460 What we claim as our invention is:

1. A method of generating at least one suggested search criterion that improves searching in a database of documents, said method comprising of:

analyzing the documents comprising the result of the first search to find at least one potential search criterion met by at least one of said documents;

465 choosing at least one search criterion among said potential search criteria that is met by a number of said documents, where said number is greater than a certain lower threshold and less than a certain upper threshold;

choosing a subset of said chosen potential search criteria such that a criterion outside the subset is met by a set of documents close to the set of documents met by
470 at least one of the search criteria in the said subset.

2. The method of claim 1, wherein said thresholds are set at some fixed percentages of the number of documents in the said first search result.

3. The method of claim 1, wherein said thresholds are adjusted based on the analysis of the said first search result.

475 4. The method of claim 1, wherein said database of documents is World Wide Web, or a subset thereof.

5. The method of claim 1, wherein types of said documents include but are not limited to hypertext documents, Web pages, text documents.

6. The method of claim 1, wherein said result of first search is a set all documents
480 meeting the search criteria of said first search.

7. The method of claim 1, wherein said result of first search is a subset of all documents meeting the search criteria of said first search.

8. The method of claim 1, where said choosing a subset is achieved by grouping said chosen search criteria into at least one group where said criteria within each said group
485 are similar with respect to searching.
9. The method of claim 8, wherein said similarity is calculated as correlation of two functions that describe the appearance of said criteria in said documents.
10. The method of claim 8, further comprising selecting at least one search criterion within at least one group and assigning said selected search criterion as representative of
490 this group.
11. The method of claim 10, wherein said selection is based in part on correlation of potential representative to other criteria within said group.
12. The method of claim 10, wherein said selection is based in part on correlation of potential representative to criteria outside said group.
- 495 13. The method of claim 10, wherein said selection is based in part on pattern of occurrences of said potential representative in said documents.
14. The method of claim 10, wherein said selection is based in part on ability of said potential representative to divide the search space.
15. The method of claim 10, wherein said selection is based in part on linguistic
500 information.
16. The method of claim 10, wherein said selection is based in part on contextual information.
17. The method of claim 1, wherein said potential criteria comprise a phrase and a procedure of matching said phrase to phrases contained in said documents.
- 505 18. The method of claim 17, wherein said phrases comprise sequences of two or more words.

19. The method of claim 17, wherein said procedure comprising matching each word in said phrase to a sequence of words in said documents.

20. The method of claim 19, wherein said procedure includes disregarding:

510 auxiliary words;

hypertext markup;

scripts; and

other information not directly related to the semantics of the document.

21. The method of claim 19, wherein said matching includes linguistically

515 normalizing word forms.

22. An interactive method of searching a database of documents comprising the following steps:

accepting the first search request from user;

executing the said search request;

520 analyzing the result of said search request execution;

calculating at least one new search criterion based on said analysis;

allowing the user to select at least one said new criteria; and

iterating said algorithm to refine the search results, wherein each subsequent iteration involves new analysis of results obtained in the previous iteration.

525 23. The method of claim 22, wherein said database of documents is World Wide Web, or a subset thereof.

24. The method of claim 22, wherein types of said documents include but are not limited to hypertext documents, Web pages, text documents.

25. The method of claim 22, wherein user can choose at least one of the said selected
530 new criteria to be added to said search request for subsequent iterations.
26. The method of claim 22, wherein user can choose a complement of at least one of
the said selected new criteria to be added to said search request for subsequent iterations,,
where complement of a criterion is defined as a new criterion that is met by a document if
and only if the said document does not meet the original criterion.
- 535 27. The method of claim 22, wherein at least one of search criteria in the said search
request can be ignored.
28. The method of claim 22, wherein at least one search engine is used to execute the
said search request.
29. A computer program product for use in a computer system, the computer program
540 product for assisting the user in searching, the computer program product comprising one
or more computer-readable media having stored thereon computer executable instructions
that, when executed by a processor, cause the computer system to perform the following:
- accept the first search request from user;
 - execute the said search request;
 - 545 present the user with the result of said search request execution;
 - analyze the result of said search request execution;
 - present user with suggested search criteria that are selected based on said analysis
to optimize the next search iteration;
 - allow user to select at least one said new criteria and add it to the search request;
 - 550 allow user to select at least one said new criteria and add its complement to the
search request; and
 - allow user to iterate the algorithm outlined here to refine the search results.